

There Will Be Numbers

Andrew Piper

05.23.16

*Journal ISSN: 2371-4549**DOI: 10.22148/16.006*

Beginnings are always hard to trace. They tend to belong more to the realm of myth, as Tristram Shandy well knew. At what point did it become necessary, in the sense of unavoidable, to use computation to study culture? Was it a certain polemic, new kinds of data (Google Books, Project Gutenberg), the rise of analytical techniques (natural language processing, machine learning), technologies such as the internet or social media, or simply that powerful social actor called "critical mass"? It is hard to say with much certainty, although I suspect people will be battling over this for years to come. For many, of course, there is nothing necessary about this approach at all. It seems profoundly *unnecessary*. It consumes resources, it is politically and technologically expedient, i.e. it fails to *resist*.

There have by now been so many polemics written for and against the use of data to study literature, culture, media and history that to offer one more rationale seems perilously unnecessary. What is needed, for sure, is more research - more research into why exactly, why right now, the computational study of culture *is* necessary. Why is it not fashion or fad, but a necessary contribution to how we understand culture? And how might we describe this need, and describe it with the circumscription and openness that is essential for building any new discipline? For surely this is what we are doing. Gone are the days of interdisciplinarity as the art of dabbling - a little linguistics or a dash of anthropology thrown in with just a touch of science studies to give that old interpretation a new shine. Here we are talking about the merger of two entire fields - of "faculties" in the early modern sense denoting a shared sense of expert ability - the incorporation of two very different critical communities, two very different forms of scholarly habitus. Computation plus culture. In an ideal world, it will equal more than the sum of its parts.

It would be wrong, and intellectually limiting, to see this undertaking solely as computer science applied to culture. Cultural analytics requires a wholesale rethinking of *both* of these categories. Computation forces us to rethink our current disciplinary practices in the humanities from the ground up. What counts as evidence? What is the relationship between theory and practice? How do we account for the technological mediations of our critique? But culture too impinges upon computation. It challenges the universalism and the neutrality implicit in many computational applications. It reminds us that knowledge is always situated, somewhere, at some time, by someone. Putting culture into computation cautions us to remember where we are when we think we know something.

It is in this spirit of mutual influence that I offer four "gaps" or problems that have retroactively come into view in the study of culture through computation and that computation is well-suited to address. These gaps can be loosely thought of as primary emphases of this journal, of what we hope CA can uniquely contribute to the landscape of research today. These are by no means meant to be exhaustive. They focus far less on the what, on content, the analysis of this or that subject, and far more on the how, on method: how computation will change the study of culture, and in turn how culture will change the study of computation.

From the anthropologist's notebook, to the historian's archive, to the media theoretician's screen or platform, to the art historian's collection, to the literary historian's book, computation has shown us how the traditional media of cultural studies are all, in their own way, flawed portals to understanding culture when used as the primary means of study. As I will discuss in the next section, none of them can by themselves solve the problem of generalization, of how to move from part to whole. In this sense, we can think of computation as cultural studies' *pharmakon*, as Plato intended it - a remedy that can also be a poison, in this case a poison to the sense of well-being founded on a widely-held belief in a discipline's prior methodological efficacy. And yet, it is also not a stretch to see culture as computation's own pharmaceutical - inserting culture into the study

of computation will allow for greater critical engagement with the ways in which computation structures meaning. Far from the medium of media, computation in the field of cultural analytics becomes part of the structure of cultural knowledge more generally. It inserts itself into an ongoing process, one actor among others. In rendering the practice of computation more legible, it opens it up to greater critical engagement.

The Evidence Gap, or the New Generality

Writing in exile in the 1940s, the German Romance philologist, Erich Auerbach, produced one of the landmark works of twentieth-century literary criticism. *Mimesis: The Representation of Reality in Western Literature* was monumental both in its scope - it covered texts from Homer to Virginia Woolf - and its categorical claims. It had as its focus "mimesis," "reality," and all of "Western Literature." It served as a potent warning to all aspiring philologists, like that terrible inscription that hung over the gates of hell in Dante's *Inferno*, not incidentally one of the books in Auerbach's massive oeuvre: "Abandon all hope, ye who enter here." Who could ever claim to possess such erudition?

Much recent work has been done to debunk Auerbach's self-portrait as an isolated exile. Istanbul was hardly the cultural backwater that he made it out to be (Auerbach had after all last taught in Marburg, population ca. 30,000). But no one, to my knowledge, has ever tried to debunk Auerbach's erudition, his immense learnedness. Who would ever presume to have read more than Auerbach? But what if he actually hadn't read *enough*?

Auerbach's work offers a cautionary tale for the future of cultural studies, but not because of its feigned orientalism, the way it establishes a too-convenient binary between East and West. Rather, its importance lies in the way it more universally inscribed exile at the very heart of cultural criticism. *Mimesis* dramatized, perhaps more than any other work in the field before or since, the crisis of metonymy as the fundamental structure of cultural criticism, the incommensurable relationship between part and whole. Every chapter begins with a passage from a "great work," where the singular instance of text is meant to stand for the singular nature of the work under review, which in turn stands for all of "occidental literature." Behind all of this lay reality itself. The cultural critic, stationed at a distance, captured the inaccessibility of the cultural whole to which his knowledge aspired. "You grasp the spirit that you resemble," says the Earth Spirit to Faust just prior to disappearing, as the whole slips once again through his fingers. This from one of the great works of literature about the limitations of humanist knowledge. It is telling that Auerbach didn't include it in his treatise.

Ever since Auerbach, we have been replaying this metonymical tragedy in the diverse fields of cultural criticism. We continue, as the saying goes, to say more and more about less and less: i.e., our parts, which do not change in scale, are becoming ever smaller in relation to the whole, which is changing in scale. A few long poems written in English stand in for "modernism," a few TV shows stretching over decades stand for the medicalization of "television," a few websites or user interviews show us "participatory culture" at work. All insightful analyses to be sure. But are we sure these examples adequately stand for the whole that they invoke? If so, how? The point is not to single out any one study or discipline or theoretical school, but merely to point out that absent computation *all* of these studies have a fundamental limitation. They are all exiled from an understanding of the representativeness of their own evidence.

Despite its title, *Mimesis* was importantly not about the "representation" of reality, but what Auerbach called "represented reality" [*dargestellte Wirklichkeit*]. The translation leads us slightly astray. There were not two separate categories (representation and reality) for Auerbach, but a single unified one. Reality was always mediated through some kind of construction, one that was historically varying in nature. This was one of the profound insights of Auerbach's work. And yet even *Mimesis*' insistence on the representedness of reality was still blind to the *representativeness* of its own examples. Whether it was the passages that stood for the works from which they were excerpted or the works that stood for the culture from which they were drawn, there was no way for Auerbach to address the fraught relationship between part and whole. How well did these local examples account for the great works from which they came or the even more general phenomenon of Occidental literature? How would it be possible to *generalize* under such terms?

Some might respond that we should simply abandon generalization and go back to what we do best: studying the particulars of cultural practice. "The task of understanding then," write Stephen Greenblatt and Catherine Gallagher in their manual *Practicing New Historicism*, "depends not on the extraction of an abstract set of principles, and still less on the application of a theoretical model, but rather an encounter with the singular, the specific, and the individual" (6). Cultural criticism, so this version goes, is at its best when it is attune to the particulars, the profound singularity, of any cultural document or

practice. We are freed from the impossible burden of reading everything because all that matters is the power of the local insight.

And yet the temptation to generalize, to scale-up the nature of one's argument, is ever-present, and it should be added, entirely legitimate. Generalization is one of the ways through our critiques assume greater social significance. In one of the great paradoxes of intellectual history, it was precisely New Historicism that drew our attention to a document's "social energy," the complex ways in which it was imbedded in a variety of social practices, all the while defending the value of the anecdotal nature of evidence ("the method of the Luminous Detail" in Ezra Pound's words). Decades worth of critical theory have shown us the value of thinking more generally about culture and the extent to which recourse to "particularity" serves as a convenient fiction, one with very clear political overtones. It is the meme of the heroic individual, like Auerbach in exile, freed from the constraints of context. But how could this ever be possible? What are the interests behind the invocation of such difference? Why do we forget our implicatedness at the very moment it matters most?

This then is one of the major contributions, and challenges, of cultural analytics. Rather than abandon generalization, the task before us is to reflect not simply on the acts of cultural representation, of Auerbach's notion of "represented reality," but on the representativeness of our own evidence, to reconstruct as contingently as possible the whole about which we are speaking. Instead of embodying the whole like the cultural critic - having read the entire archive or seen all the images - the cultural analyst focuses instead on the act of construction itself. The cultural analyst is self-conscious about being implicated in the knowledge that is being created.

Whereas the social sciences would normally speak of things like "samples" and "bias" in this context, I prefer the notion of *representativeness*. First, it aligns well with the attention in cultural criticism with "representation," that is, with symbolic forms more generally. What makes cultural criticism unique is its attention to the constructedness of culture, to the making, creativity, and communication that surround human cultures, rather than their laws or essences. Second, it captures the constructed nature of knowledge itself, the necessary incompleteness that surrounds any part in relationship to a larger whole. Culture, like cultural criticism, is never finished. Terms like "sample" and "bias" presuppose some stable, knowable whole to which there is presumably a correct representation. "Representativeness," on the other hand, simply says that every part is always a representation. It values the curation of data over the amount of data. We are not on the outside, as in Auerbach's model of the European in Oriental exile, rationally constructing things in a disinterested way from a distance. Rather, we are always crucially inside of knowledge, representing our representations of culture from within.

The Theory Gap, or the New Explicitness

Most readers will be familiar with the so-called theory wars, with the various charges of too much or too little theory that have surrounded the various fields of cultural study. What has been missing through all of this, however, is a reflection on how we move from theory to practice. The gap in theory is not a gap *of* theory, of too little, but a gap *from* theory, of what comes next.

The problem should be familiar to anyone who has ever read an article in the humanities. To take one example: "The essays collected here examine the present moment of digital anachronism - wherein the most contemporary of media are haunted by celluloid spectres - through a diverse series of images, texts, concepts spanning film history." The question that immediately arises is how? What does one mean by "digital anachronism" or "celluloid spectres" or "film history"? Where are they going to find these things and how will they know how much of it they have found or whether there is more of it elsewhere or whether other things are more important to the objects they're talking about?

The practice of cultural analytics fill in these gaps. It says exactly how many documents are under consideration and which ones. It defines what exactly is meant by loaded terms such as "haunted," "power," "diversity," "diegesis," "race," "genre" or "film history." When I talk about spatiality in contemporary television, I mean the use of interior or exterior scene cues in 1,182 screenplays. When I talk about "conversion" in the novel I mean a distinction in the heterogeneity of speech between the first and second halves of novels written over a century and a half and three languages. Then I define the formula for how I calculate "heterogeneity of speech."

There is an aspect of reductiveness to all of this that is understandably hard for many to accept. For cultural critics, things are always more complex than they seem. That may in fact be one of the most elementary definitions of "culture" - a

practice or artifact that conveys more than one meaning, that has at its base a commitment to something more than self-evidence. But such reductiveness is the cost of scale and scale is the price of generalization. We cannot know something at the general level as complexly as we can at the local level. There is an inverse relationship between the number of things considered and the complexity of what can be known about them. In this sense, complexity in cultural analytics is not an inherent property of the documents themselves, but an emergent property of the documents taken together. It is produced through the aggregation of lower-level phenomena.

Sometimes, nothing happens. This is called a negative result and is as important as the novel insight of something previously unseen. It is why there is so much more chance in cultural analytics than in traditional cultural studies. In the normal procedures of cultural studies, the researcher's ideas evolve as s/he comes into contact with more and more documents. But only certain documents are ever encountered and from these an even smaller number are chosen to be discussed and these are chosen because they confirm the researcher's beliefs. We *never* hear of what didn't fit or how the idea under review wasn't actually very prevalent in the vast trove of things being explored. The cultural analyst, on the other hand, undergoes the same process of exploration and refinement in developing a hypothesis, but then tests this hypothesis against a much greater number of documents. One can guess what might happen at these larger levels of scale - that's why it's a hypothesis - but one can never know for sure in advance. Unlike the imperious pronouncements of the cultural critic, there is an element of error, of uncertainty, built into the task of the cultural analyst.

There is also an element of disenchantment too to all of this that needs to be taken into consideration. Such methods push against the black-box of charisma and insight that have been the historical foundations of knowledge in the humanities - and by extension the elitism at their core. How did a critic come to arrive at this insight? What did he or she read or see prior to its genesis? What were the procedures through which the mind conjured such brilliance? Genius!

The cultural analyst by contrast lays bare as much as possible of this intellectual process. In this, the practice aims to be more democratic, not less so as many claim. Algorithms require that we define every step of our intellectual process. In their simplest sense, they are nothing more than a set of instructions. Humanists have traditionally been opaque about their methods - about what evidence has been considered and what is considered as evidence. Mediating knowledge through computation is not a dehumanization of the "humanities" or "culture." It is an act of mirroring, one where we show ourselves the cognitive steps we are taking to arrive at an argument about culture. Computation in this sense isn't fast - it slows us down and forces us to be more self-reflective.

One of the core values of such explicitness is that it allows for a model of knowledge creation that is more consensus-driven than has been the case in the past. Cultural criticism in its traditional form resembles an agonistic process - one continually overturns the views of others in a process seemingly without end. There is nothing to mediate the dispute. As a cultural critic, one feels like the child of parents who argue incessantly out of a sense of sport or even boredom, all in the name of a higher principle. The explicitness of cultural analytics means that others can share in the steps of the analyst's knowledge. They can correct those steps and challenge them, or they can build on them and refine them because those steps have been made more legible. In either case, the tools and information for mediating disagreement are made more mutually available. Cultural analytics makes the study of culture more architectonic rather than agonistic, more social and collective. There is a basic politics to this practice that has largely been overlooked.

The Self-Reflexive Gap, or the New Recursivity

As I've already emphasized, there is a self-reflexive quality to cultural analytics that is to my mind one of its most important features. In being more explicit, in documenting and theorizing our practices more extensively, the cultural analyst becomes more aware of hidden assumptions and buried beliefs. In self-consciously drawing lines around data, the cultural analyst marks out the terrain of what one knows (and thus does not know). Far from a space of universal knowledge, cultural analytics marks out more clearly the contingencies of knowledge. Unlike the cultural critic who pronounces on the whole from an extraordinarily small selection (à la Auerbach), the cultural analyst circumscribes the space of knowledge. There is an epistemological humility that resides at the core of cultural analytics.

One of the key concepts operative in computational research that has so far been missing from traditional studies of culture is that of modeling. A model is a metonymical tool - a miniature that represents a larger whole. But it is also recursive in that it can be modified in relationship to its "fit," how well it represents this whole. There is a great deal of literature on

the role of modeling in knowledge creation and this should become core reading for anyone undertaking cultural analytics. The more we think about our methods as *models* the further we will move from the confident claims of empiricism to the contingent ones of representation. Under certain conditions, it is true that (i.e. replicable and stable)... Modeling puts computation not on the outside of what is known but as part of the process itself. Cultural analysts are not actors outside of culture, but the very things they produce become part of culture.

The best models of cultural analytics will thus take into account these recursive structures of cultural knowledge. These can include the study of disciplinary structures or media infrastructures, where we build into our models the epistemological and technological orientations of a given discipline, identifying the ways they influence what we know. Culture does not come to us outside of these spaces of reception. Or one could include the study of computational models themselves, showing the extent to which different algorithms generate different kinds of meaning. This is what I meant earlier about the ways in which culture impacts computation - computation is not on the outside of culture looking in, but part of the process through which culture is generated. To rehash an old philosophical slogan with a computational twist, in cultural analytics there is no outside of culture.

The Relevance Gap, or the New Impactfulness

This brings me to my fourth and final gap, and arguably the one that will be the most polemical for researchers in the humanities. Few words have a more distasteful ring about them today than the word "impact." Impact is akin to selling out. It undermines that Auerbachian stance at a distance, the independence of the researcher's critical gaze, the autonomy rather than heteronomy that is at the core of critical being. But as should be clear by now, cultural analytics argues that there is no comfortable outside from which one can neutrally observe culture; there is no space where one is not implicated. Computation unquestionably inserts the analyst into the institutional ligatures that make research possible. In adapting to the evidentiary and institutional paradigms of contemporary knowledge, it rethinks critique not as irrevocably outside of something but implicated in it and thereby potentially transformative. It attempts to align research more closely with the social tensions in which that research is being produced and to bring such research to bear on those debates. The qualities of explicitness and reductiveness that surround computational modeling can paradoxically serve as the very conditions of enabling arguments to circulate more widely and, potentially, to create social change. As we have already begun to see, the work of cultural analytics can facilitate the practice of cultural advocacy.

Welcome to CA

So what kinds of articles is CA hoping to promote? The simple answer is articles that address the challenges I have outlined above across a diverse set of cultural domains, spaces, times, practices and media. We are looking for articles that address head on the representativeness of their evidence; that make explicit not simply their theoretical terms, but the movement from theory to implementation; that reflect on the extent to which the process of modeling has helped construct the very things that one has found; and finally, that attempt to address and intervene in contemporary debates. No one article will do all these things in equal measure. Some articles may want to test different models of representativeness. Others may want to explore the extent to which different models produce different interpretations of the same object. Still others may wish to bring existing techniques to bear on questions of cultural inequity. In each instance - and here I offer one final gap that is specific to the pre-history of this emerging field - what is fundamental to all of these exercises is the presence of an *argument*. What precisely does computation allow one to claim that has not been seen before or that was uncertain in the world of anecdotalism? For all the novelty surrounding this undertaking, what is constant is the imbeddedness of research within existing paradigms and debates and the need to engage critically with these traditions. Cultural analytics emphatically belongs to the longer tradition of cultural studies.

There are already plenty of examples of research of this nature out there. I won't list them here (partially because I also want us to reimagine the highly partisan nature of the "literature review"). The goal of CA is to bring this work under one shared intellectual space - to foster dialogue, experimentation, and self-reflection in this novel undertaking. As our first issue tries to show, in some cases the analysis might be of a new kind of cultural object like applause (Clement et al) or an old object like photography understood in a much broader context (Manovich). In others, it might be a new way of measuring the

social coherence of a category like "genre" (Underwood) or the relationship between genres and how these kinds of writing intersect (Vierthaler). Or it might be an attempt to construct new kinds of disciplinary histories (Manovich). In the future we aim to have articles that review humanities data sets as well, reflecting on the infrastructure of cultural knowledge and how those acts of curation influence work in the field. And we will have a "debates" section where we try to promote pieces that take a stand on theoretical issues important to the field today. Finally, we intend to run our own analytics, to subject CA to the same self-critique as the field as a whole.

Cultural analytics thus encompasses the variety of human (and non-human) cultures and media that have gotten us to where we are today. What unites it is a belief that computation can show us things about culture that previous media and their metonymical impasses could not. The proof, as they say, is in the pudding.

Andrew Piper, McGill University

